Title

Genome-scale analysis identifies paralog lethality as a vulnerability of chromosome 1p loss in cancer.

Journal

Nature genetics, 50(7)

ISSN

1061-4036

Authors

Viswanathan, Srinivas R
Nogueira, Marina F
Buss, Colin G
et al.

Publication Date

2018-07-01

DOI

10.1038/s41588-018-0155-3

Peer reviewed

HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2018 December 28.

Published in final edited form as:

Nat Genet. 2018 July ; 50(7): 937–943. doi:10.1038/s41588-018-0155-3.

Genome-scale analysis identifies paralog lethality as a vulnerability of chromosome 1p loss in cancer

Srinivas R. Viswanathan[1,2,3], **Marina F. Nogueira**[#1,2], **Colin G. Buss**[#4,5], **John M. Krill-Burger**[2], **Mathias J. Wawer**[6], **Edyta Malolepsza**[2,11], **Ashton C. Berger**[1,2], **Peter S. Choi**[1,2,3], **Juliann Shih**[2], **Alison M. Taylor**[1,2,3], **Benjamin Tanenbaum**[2], **Chandra Sekhar Pedamallu**[1], **Andrew D. Cherniack**[2], **Pablo Tamayo**[2,7], **Craig A. Strathdee**[2], **Kasper Lage**[2,11], **Steven A. Carr**[2], **Monica Schenone**[2], **Sangeeta N. Bhatia**[2,3,4,5,8,9,10], **Francisca Vazquez**[2], **Aviad Tsherniak**[2], **William C. Hahn**[1,2,3], and **Matthew Meyerson**[1,2,3,♦]

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA

[2]Broad Institute of MIT and Harvard, Cambridge, MA, USA

[3]Harvard Medical School, Boston, MA, USA

[4]Harvard-MIT Department of Health Sciences and Technology, Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Boston, MA USA

[5]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA.

[6]Chemical Biology and Therapeutics Science Program, Broad Institute of Harvard and MIT, Cambridge, MA, USA

[7]UCSD Moores Cancer Center and Department of Medicine, University of California, San Diego, La Jolla, California

[8]Howard Hughes Medical Institute, Chevy Chase, MD, USA

[9]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA.

♦To whom correspondence should be addressed: Matthew Meyerson, Dana-Farber Cancer Institute, Boston, MA 02115, Phone: 617-632-4768, matthew_meyerson@dfci.harvard.edu.

AUTHOR CONTRIBUTIONS:

S.R.V. and M.M. conceived of the research and wrote the manuscript. S.R.V. and M.F.N. performed experiments. C.G.B. performed *in vivo* xenograft experiments. M.J.W., S.R.V., P.S.C. performed data analysis on RNA sequencing data. J.M.K-B. and S.R.V. performed data analysis on shRNA and CRISPR screening data. A.C.B., A.M.T. and J.S. performed copy number analysis on TCGA and cell line data. C.A.S. assisted in generation of shRNA reagents. P.T., A.D.C., and C.S.P performed or oversaw data analysis. B.T., K.L., S.A.C., E.M, and M.S. performed mass spectrometry or were involved in downstream data analysis. S.N.B. oversaw *in* vivo xenograft experiments. F.V., A.T., and W.C.H. directed shRNA and CRISPR screening efforts. M.M. directed the overall project.

COMPETING FINANCIAL INTERESTS:

A.C.B., A.D.C., C.A.S. and M.M. receive research support from Bayer Pharmaceuticals.
M.M. is a scientific advisory board member, consultant for, and holds equity in OrigiMed.

DATA AVAILABILITY STATEMENT

The original mass spectra may be downloaded from MassIVE (see URLs) using the identifier: MSV000082292. RNA-Seq data can be accessed at NCBI Gene Expression Omnibus (GSE113848) (see URLs). Code for analysis of IP-MS data is deposited in the Lage Lab website (see URLs). The authors declare that other data supporting the findings of this study are available within the paper and its supplementary information files. Other source data are available from the corresponding author upon reasonable request.

[10]Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA

[11]Department of Surgery, Massachusetts General Hospital, Boston, Massachusetts, USA.

[#] These authors contributed equally to this work.

INTRODUCTORY PARAGRAPH

Functional redundancy shared by paralog genes may afford protection against genetic perturbations, but it can also result in genetic vulnerabilities due to mutual interdependency[1–5]. Here, we surveyed genome-scale shRNA and CRISPR screening data on hundreds of cancer cell lines and identified *MAGOH* and *MAGOHB*, core members of the splicing-dependent exon junction complex, as top-ranked paralog dependencies[6–8]. *MAGOHB* is the top gene dependency in cells with hemizygous *MAGOH* deletion, a pervasive genetic event that frequently occurs due to chromosome 1p loss. Inhibition of *MAGOHB* in a *MAGOH*-deleted context compromises viability by globally perturbing alternative splicing and RNA surveillance. Dependency on *IPO13*, an importin-β receptor that mediates nuclear import of the MAGOH/B-Y14 heterodimer[9], is highly correlated to dependency on both *MAGOH* and *MAGOHB*. Both *MAGOHB* and *IPO13* represent dependencies in murine xenografts with hemizygous *MAGOH* deletion. Our results identify *MAGOH* and *MAGOHB* as reciprocal paralog dependencies across cancer types and suggest a rationale for targeting the *MAGOHB-IPO13* axis in cancers with chromosome 1p deletion.

The systematic integration of data from genomic characterization and genetic screening of cancer cell lines can identify gene dependencies induced by specific somatic alterations and inform the development of targeted therapeutics. For example, several studies have revealed that inactivation of specific driver or passenger genes may confer dependency on functionally redundant paralogs[2,3,10–13]. Paralog dependencies have also emerged as important targets in recent genome-scale functional genomic screens[4,5], underscoring the importance of further characterizing this class of cancer vulnerabilities.

To systematically identify paralog dependencies that may represent attractive cancer targets, we analyzed data from pooled, genome-scale short hairpin RNA (shRNA) screening of 501 cancer cell lines[5,14]. We determined the correlation between a dependency on a gene[5] and loss of function of its paralog across 10,287 paralog pairs (Supplementary Figure 1; Supplementary Note). We identified 167 genes for which dependency was significantly correlated to loss of a paralog (1.6% of paralog test pairs at $q < 0.05$), including many previously reported paralog dependencies (e.g. *ARID1B* dependency with *ARID1A* inactivation[10], *SMARCA2* dependency with *SMARCA4* inactivation[11], *UBC* dependency with *UBB* inactivation[5], and *FERMT1* dependency with *FERMT2* inactivation[5]). However, of these 167 paralog dependency pairs, only 7 were "symmetric", in which dependency for each of the genes in the pair was significantly correlated to inactivation of its partner paralog (Fig. 1a-b; Supplementary Table 1). A similar analysis of data from genome-scale CRISPR screening of 341 cell lines[15] identified 125 significant paralog dependencies (1.4% of paralog test pairs at $q < 0.05$), of which 7 pairs were symmetric (Supplementary Table 2; Supplementary Note). Paralog genes arise via ancestral duplication events and may functionally diverge over time[1,16]. Symmetric paralog pairs likely share complete functional

redundancy, making them particularly attractive targets for "collateral lethality" strategies[2]. An enrichment for RNA-splicing related genes was noted among symmetric, but not asymmetric, paralog pairs in the shRNA and CRISPR screening datasets (Supplementary Table 3), suggesting that redundant essentiality may be exploited to target splicing-related pathways.

One symmetric paralog pair was shared between the shRNA and CRISPR datasets: *MAGOH-MAGOHB*; a second pair, *FUBP1-KHSRP*, was highly significant for symmetry in the shRNA data and borderline significant in the CRISPR dataset (q_1=0.0547) (Fig. 1a-1b; Supplementary Figure 1; Supplementary Tables 1-2)[15]. We focus here on validation of the former pair. *MAGOH* and *MAGOHB* encode core members of the exon-junction complex (EJC), a multiprotein complex that is deposited on mRNAs at the time of splicing and that mediates diverse downstream processes including mRNA transport, stability, and nonsense mediated decay (NMD)[6,17].

Using both shRNA and CRISPR technologies, we individually validated *MAGOHB* dependency in the setting of *MAGOH* loss, as well as *MAGOH* dependency in the setting of *MAGOHB* loss. Furthermore, in a cell line without hemizygous deletion of either paralog, knockdown of either *MAGOH* or *MAGOHB* individually was tolerated, but the combination was lethal (Supplementary Figure 2). We noted that *MAGOHB* dependency in the setting of *MAGOH* inactivation was particularly pronounced based on: 1) effect size (log-fold difference in *MAGOHB* dependency between *MAGOH*-inactivated and non-*MAGOH*-inactivated cell lines) and 2) *MAGOHB* scoring as a robust 6σ differential dependency (having a dependency score in some cell lines greater than six standard deviations below its mean dependency score across all cell lines) in both the RNAi and CRISPR screening data. We therefore sought to further characterize *MAGOHB* dependency in the setting of *MAGOH* loss.

MAGOHB was the top differential dependency in cells with hemizygous deletion of *MAGOH* (Fig. 1c; Supplementary Tables S4-S5; Supplementary Note) and dependency on *MAGOHB* was predicted by low expression of *MAGOH,* consistent with the notion that hemizygous deletion of *MAGOH* leads to its decreased expression (Supplementary Figure 3). shRNA-mediated knockdown of *MAGOHB* led to a decrease in cell viability and colony forming capacity in three *MAGOH*-deleted cell lines, but not in control cell lines euploid for *MAGOH* (Fig. 1d-e; Supplementary Figure 4). Ectopic expression of MAGOH in a *MAGOH*-deleted cell line fully rescued *MAGOHB* dependency, indicating that *MAGOHB* dependency in *MAGOH*-deleted cells is solely due to *MAGOH* loss, and consistent with complete functional redundancy between these paralogs[8] (Fig. 1f; Supplementary Figure 4). CRISPR/Cas9-mediated deletion of *MAGOH* in a cell line with two copies of *MAGOH* also conferred *MAGOHB* dependency (Supplementary Figures S5-S6).

To assess the clinical contexts in which these dependencies might be exploited, we next surveyed the frequency of *MAGOH* and *MAGOHB* loss in tumor cohorts from the Cancer Genome Atlas (TCGA). We observed pervasive hemizygous *MAGOH* loss across tumor types (frequency of 21% (1675/8009) in the entire TCGA dataset, and >50% in multiple tumor types). Moreover, *MAGOH* deletion most frequently occurs as a result of arm level

deletion of chromosome 1p across human tumors (Fig. 1g; Supplementary Table 6). We confirmed that chromosome 1p-deletion status correlates with *MAGOHB* dependency in the genome-scale CRISPR screening data (Supplementary Figure 7). In the context of neuroblastoma – where 1p deletion is a hallmark event in a subset of tumors[18] – *MAGOHB* knockdown was lethal in a 1p-deleted, but not a 1p-neutral, cell line (Supplementary Figure 7). *MAGOHB* is located on chromosome 12p, an arm also recurrently lost across tumor types, albeit with markedly lower frequency than chromosome 1p (Supplementary Figure 8). Analysis of genome-scale CRISPR screening data confirmed a reciprocal dependency on *MAGOH* in the setting of chromosome 12p deletion. Interestingly, we also observed mutual exclusivity between chromosome 1p and chromosome 12p co-deletion in many tumor types, suggesting that concurrent loss of both *MAGOH* and *MAGOHB* may be poorly tolerated (Supplementary Figure 8). We conclude that *MAGOH* and *MAGOHB* represent potential vulnerabilities in large, genetically defined, subsets of tumors.

MAGOH and MAGOHB constitute core components of the EJC[8]; EJC deposition at exon-exon junctions allows transcripts containing premature termination codons (PTCs) to be identified and targeted for degradation via NMD[6,17]. We therefore hypothesized that *MAGOHB* inhibition in the setting of decreased *MAGOH* dosage may compromise cell viability by perturbing RNA splicing and RNA surveillance. To evaluate the global transcriptomic consequences of *MAGOHB* inhibition, we performed RNA sequencing on hemizygous *MAGOH*-deleted ChagoK1 cells in the presence or absence of *MAGOHB* knockdown, with or without ectopic re-expression of MAGOH. We observed an increased expression of NMD biotype transcripts upon *MAGOHB* knockdown in ChagoK1 cells (Fig. 2a, left). In contrast, *MAGOHB* knockdown in MAGOH-reconstituted ChagoK1 cells was well-tolerated without a notable shift in NMD biotype transcript distribution (Fig. 2a, right). We next sought to determine whether the upregulation of NMD isoforms upon *MAGOHB* knockdown in ChagoK1 cells was occurring at the expense of other transcript biotypes. Among genes that had significantly upregulated NMD isoform(s) upon *MAGOHB* knockdown, we observed a significant proportional decrease in coding isoform expression in ChagoK1 cells but not MAGOH-reconstituted ChagoK1 cells (Fig. 2b, compare left and right). To investigate whether particular splice event classes were driving this redistribution of isoform types, we quantified the proportion of differentially spliced events of each class that were more common in either the absence (Fig. 2c, red) or presence (Fig. 2c, blue) of *MAGOHB* knockdown in either ChagoK1 cells or MAGOH-reconstituted ChagoK1 cells. As compared with *MAGOHB* knockdown in MAGOH-reconstituted ChagoK1 cells, *MAGOHB* knockdown in ChagoK1 cells resulted in reduced cassette exon inclusion and increased intron retention (Fig. 2c). Therefore, many global transcriptomic effects of *MAGOH/B* insufficiency appear attributable to alterations in these two splice event types, indicative of a defect in exon definition/recognition.

We identified 22 instances in which there was both a significant absolute upregulation of an NMD isoform (beta > 1 in differential expression analysis using Kallisto[19]) and corresponding downregulation of at least one protein coding isoform (beta < −1) (Supplementary Table 7). These genes were significantly enriched for pathways involved in mRNA splicing and mRNA processing (Supplementary Table 8; Fig. 2d). Intriguingly, among the seven splicing-related genes driving this enrichment were four genes (*SRSF2,*

SRSF7, HNRNPDL, HNRNPH1) reported to auto-regulate their expression via alternative splicing-nonsense mediated decay (AS-NMD) loops[20,21]. Such AS-NMD loops, many of which are in splicing-related genes, involve ultraconserved, regulated alternative splicing events that induce NMD-substrates, thus maintaining homeostatic control of gene expression[20,21]. We observed perturbations in isoform distributions of *HNRNPDL* and other splicing-related genes upon *MAGOHB* knockdown in ChagoK1 cells, but not in MAGOH-reconstituted ChagoK1 cells (Fig. 2e; Supplementary Figures 9–10; Supplementary Note). Altered RNA isoform abundance accompanied by changes in the levels of functional protein, either via disruption of AS-NMD loops or through other mechanisms, could have deleterious direct and indirect consequences on cellular splicing.

Given these transcriptomic consequences of MAGOH/MAGOHB insufficiency, we next sought to determine whether *MAGOH* loss unveils a broader dependency on splicing/NMD-related complexes. We performed immunoprecipitation and mass spectrometry to identify MAGOH- and MAGOHB-associated binding partners and found that these interactors, which include many splicing-related genes, were enriched for gene dependencies correlated to both *MAGOH* and *MAGOHB* dependencies. However, these dependencies were weaker than the reciprocal *MAGOH/B* paralog dependencies driven by redundant essentiality (Supplementary Figure 11; Supplementary Tables 9-11; Supplementary Note).

MAGOH and MAGOHB share near-identity at the protein level and functional and crystallographic studies do not reveal domains easily amenable to targeting by small molecules[22]. To identify other more tractable targets that might indirectly affect MAGOH/MAGOHB function, we interrogated the genome scale shRNA screening data for gene dependencies highly correlated to either *MAGOH* or *MAGOHB* dependency. Strikingly, *IPO13* emerged as the top, outlier correlated gene dependency to both *MAGOH* and *MAGOHB* (Fig. 3a; Supplementary Table 12). *IPO13* is a bidirectional karyopherin responsible for nuclear import of the MAGOH/B-Y14 heterodimer, a function critical for recycling of the EJC; it is also located on chromosome 1p in proximity to *MAGOH,* and the two genes are frequently co-deleted[17] (Fig. 3b; Supplementary Table 13).

We observed a selective *IPO13* dependency in *MAGOH*-deleted cell lines compared to non-deleted cell lines (Fig. 3c) and found that dependency on *IPO13* in *MAGOH*-deleted H460 and H1437 cells was partially attenuated by *MAGOH* re-expression (Fig. 3d; Supplementary Figure 12; Supplementary Note). Knockdown of *IPO13* in *MAGOH*-deleted cells led to cytoplasmic accumulation of MAGOH/MAGOHB and subsequent upregulation of the NMD-substrates SC1.6 and SC1.7, an effect that was rescued by MAGOH re-expression (Fig. 3e; Supplementary Figure 13). This suggests that IPO13 dependency in *MAGOH*- and *MAGOHB*-deleted cells is mediated in part by defective shuttling of *MAGOH/B*, resulting in mis-splicing and impaired RNA surveillance. Haploinsufficiency of *IPO13*, as occurs when *MAGOH* and *IPO13* are co-deleted, may also contribute to *IPO13* dependency in some contexts.

Finally, we sought to validate *MAGOHB* as a target *in vivo*. We formed xenografts from H1437 cells (which carry a hemizygous deletion in *MAGOH*) transduced with a lentiviral vector encoding a doxycycline-inducible shRNA against *MAGOHB*. Xenograft growth was

significantly impaired upon *MAGOHB* knockdown. (Fig. 4a-b). To next assess this dependency using a more therapeutically tractable system, we used tumor-penetrating nanocomplexes (TPNCs) capable of targeted delivery of siRNAs to the cytosol of tumor cells[23]. The TPNCs were decorated with the tumor-homing peptide iRGD, which allows for targeted delivery of siRNAs to tumor cells expressing surface NRP1/αvβ3; both receptors are expressed on H1437 cells (Fig 4c). H1437 xenograft growth was significantly impaired upon intra-tumoral injection of si-*MAGOHB* and si-*IPO13*-containing TPNCs, but not TPNCs containing control siRNA against *GFP* (Fig. 4d-e; Supplementary Figure 14). This finding was confirmed using a second TPNC-system using a distinct tumor-homing peptide, Lyp-1 (Supplementary Figure 14). Additionally, tumors treated with si-*MAGOHB*-containing TPNCs displayed higher levels of cleaved caspase-3, indicating that targeting *MAGOHB* in a *MAGOH*-hemizygous context triggers apoptotic cell death (Supplementary Figure 14). Thus, *MAGOHB* and *IPO13* represent potential *in vivo* targets in a *MAGOH*-deleted context, and this paralog vulnerability may be exploited by antisense or RNAi-based approaches.

Hemizygous chromosome arm loss is one of the commonest features of cancer genomes[24,25] and rational therapeutic targeting of this class of somatic events would therefore be attractive. Prior studies have identified several candidate targets unmasked by genomic loss[4,5,26,27]. Here, we integrate genomic characterization and genome-scale functional screening of cancer cell lines to systematically extend such studies. We identify a set of robust paralog dependencies that may provide the foundation for future target validation efforts and show that hemizygous loss of the *MAGOH* gene on chromosome 1p confers novel vulnerabilities on *MAGOHB* and *IPO13*, perhaps due to decreased nuclear reserve of MAGOH/MAGOHB (Supplementary Figure 15). Insufficient MAGOH/MAGOHB dosage perturbs splicing and RNA surveillance and adds to growing evidence implicating splicing as a cancer dependency[27–29]. Therapeutic approaches to targeting *MAGOH*-deleted cells may involve either direct *MAGOHB* transcript suppression (such as through antisense/RNAi approaches), targeted MAGOHB protein degradation, or indirect suppression of MAGOH/MAGOHB activity via inhibition of IPO13. Antisense/RNAi-based approaches may be well suited to the exploitation of paralog dependencies, as they may allow for selective targeting of paralogs that show greater variability on the nucleotide level than on the protein level. Targeted protein degradation approaches have also recently proven to be a promising means to target conventionally “undruggable” genes[30–33], including RNA splicing factors[34]. In the case of *IPO13* dependency, small molecule inhibitors of other importin family members have been described, raising the possibility that IPO13 can be selectively targeted using a similar strategy[35–37]. As hemizygous loss of chromosome 1p is extremely common across multiple tumor types, these or other approaches to targeting this pathway may have future biomarker-driven therapeutic applications. More broadly, our work can be generalized to cancers with other chromosome arm deletions and underscores the power of intersecting comprehensive molecular characterization and functional genomic studies of cancer cell lines.

ONLINE METHODS

Cell culture

Cell line stocks used for validation experiments were obtained either from the Cancer Cell Line Encyclopedia repository at the Broad Institute or from M.M's laboratory, with original sources being either the American Type Culture Collection, the European Collection of Authenticated Cell Cultures, the Health Science Research Resources Bank, Korean Cell Line Bank or academic laboratories. Cell line identity was verified by either STR profiling or Affymetrix SNP profiling. Cells were cultured in media specified by the source repository, supplemented with 100 IU/ml penicillin, 100 μg/ml streptomycin, 2 mM L-glutamine, and 100 μg/mL Normocin (Invivogen). Mycoplasma testing was performed in source repository prior to creation of frozen stocks and repeated periodically if lines were persistently maintained in culture.

Lentiviral constructs and transduction of cell lines

For overexpression experiments, ORFs were expressed from within the pLX304-Blast-V5 vector[38] (Addgene #25890, Blasticidin resistance) using pLX304-eGFP as an overexpression negative control. Ectopic expression of untagged MAGOH was performed either using pLX304 (with stop codon introduced prior to V5 tag) or Gateway-compatible, Hygromycin-resistant, doxycycline-inducible overexpression vector with cDNA expression driven from Tet-regulated CMV promoter, created by modification of prior similar vectors[39,40]. MAGOH-reconstitution experiments were performed with V5-tagged MAGOH with the exception of those shown in Supplemental Figure 4e, which were performed using an untagged construct. For shRNA experiments, constitutive shRNAs were expressed from the pLKO.1 vector[41] (Addgene # 10878, Puromycin resistance) using an shRNA targeting GFP (shGFP) as a negative control. shRNA constructs were obtained from the RNAi Consortium shRNA collection (see URLs) Inducible shRNAs were cloned into a Gateway-compatible doxycycline-inducible lentiviral shRNA expression system (G418 resistance), as described[39]. sgRNAs were cloned into lentiCRISPRv2 (Addgene #52961) as described (see URLs). shRNA and sgRNA target sequences are listed in Supplementary Table 14.

Lentivirus was produced in HEK293T cells as per the "low throughput viral production" protocol on the RNAi Consortium Portal (see URLs). Cells were transduced with lentivirus by spin-infection (2250 rpm for 30 minutes) in the presence of 8 μg/ml polybrene, followed by antibiotic selection beginning 24 hours thereafter. Following completion of antibiotic selection, cells were seeded for downstream assays as described.

URLs
Broad RNAi Consortium, http://www.broadinstitute.org/rnai/public
LentiCRISPRv2 Cloning Protocol, http://genome-engineering.org/gecko/wp-content/uploads/2013/12/lentiCRISPRv2-and-lentiGuide-oligo-cloning-protocol.pdf
Lentiviral Production Protocol, http://portals.broadinstitute.org/gpp/public/resources/protocols
Bash script for expectation maximization algorithm, http://www.lagelab.org/resources
HUGO Gene Nomenclature Committee, http://www.genenames.org/cgi-bin/genefamilies
rMATS2Sashimiplot, https://github.com/Xinglab/rmats2sashimiplot
ENSEMBL Biomart, http://www.ensembl.org/biomart DepMap Portal, http://depmap.org
MassIVE, http://massive.ucsd.edu
NCBI GEO, https://www.ncbi.nlm.nih.gov/geo/

Gene knockdown or knockout was confirmed by quantitative RT-PCR for *MAGOH/B* as the paralogs cannot be distinguished on Western blotting. Gene overexpression was confirmed by Western blotting.

Generation of *MAGOH*-knockout cell lines

For generation of *MAGOH*-knockout cell lines, Heya8 cells were transiently transfected with either a non-targeting guide (sgGFP) or MAGOH sgRNA expressed from within plentiCRISPRv2. Following 72 hours of selection with puromycin, the bulk resistant population was sorted at single cell density into 96 well plates using a MoFlo Astrios Cell Sorter (Beckman Coulter). Clonal cell lines were expanded and assessed for MAGOH knockout by quantitative PCR.

Western blotting

Whole-cell extracts for immunoblotting were prepared by incubating cells on ice in RIPA lysis buffer (ThermoFisher Scientific) plus protease inhibitors (cOmplete, Mini, EDTA-free, Roche) for 20 minutes. Following centrifugation (>16,000 r.c.f. for 15 minutes), protein lysates were quantitated using the Pierce BCA Protein Assay Kit (ThermoFisher Scientific). Lysates were separated by SDS-PAGE and transferred to nitrocellulose membranes using the iBlot2 system (Life Technologies). Two-color immunoblotting was performed using the LI-COR platform (LI-COR Biosciences) with IRDye 800CW and IRDye680RD secondary antibodies (mouse, IRDye 680LT Donkey anti-Mouse IgG (925–68022) used at 1:10000; rabbit, IRDye 800CW Goat anti-Rabbit IgG (926–32211) used at 1:10000). Imaging was performed on an Odyssey CLx Infrared Imaging System. Loading control and experimental protein were probed on the same membrane in all cases. For clarity, loading control is cropped and shown below experimental condition in all panels regardless of relative molecular weights of the two proteins.

Primary antibodies and dilutions used were as follows. HNRNPDL: Thermo Fisher Scientific (PA5–35896), 1:2000. Vinculin: Sigma-Aldrich (V9264), 1:4000. MAGOH: Santa Cruz Biotechnology (sc-271365), 1:250 or Abcam (ab38768) (1:500). Actin: Cell Signaling Technology (D6A8, mAb #8457), 1:1000 or Cell Signaling Technology (8H10D10, mAb #3700), 1:2000. PSPC1, Santa Cruz Biotechnology (sc-374181), 1:100. HNRNPH1, Bethyl Laboratories (A300–511A), 1:1000.

Quantitative PCR

RNA isolation was performed using the RNeasy Mini Kit (Qiagen). cDNA preparation was performed using Superscript III cDNA synthesis kit (Thermo Fisher Scientific). PCR reactions were prepared using Taqman Gene Expression Mastermix (Thermo Fisher Scientific) and PrimeTime qPCR probe-based assays (IDT) using *HPRT1* as an internal normalization control. Taqman assay ID catalog numbers are as provided in Table S3. Real-time quantitative PCR was performed on a QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems) and results were quantitated using the ΔΔCt method. For quantification of SC35 NMD substrates, quantitative PCR was performed using a Power SYBR Green PCR Master Mix (Thermo Fisher Scientific) and primer sequences for SC1.6 and SC1.7 as described[8] using *β-actin* as an internal normalization control.

Cell Viability and Colony Formation Assays

For cell viability assays, cells were seeded in 96-well plates in 100 μl of medium after lentiviral transduction and completion of antibiotic selection. For inducible hairpin experiments, equal numbers of cells were seeded for both “-Dox” and “+Dox” conditions and medium was supplemented with 100 ng/ml doxycycline in the “+Dox” condition. Cells were seeded in 96 well plate format (range 1000–8000 per well, depending on the cell line). At 7–10 days after cell seeding, cell viability was assessed using the Cell Titer-Glo luminescent cell viability assay (Promega) using either an EnVision Multilabel Reader (Perkin Elmer) or a Spectramax M5 plate reader (Molecular Devices).

For colony formation assays, cells were seeded in 12-well plates at a density of 2000–8000 cells per well after lentiviral transduction and completion of antibiotic selection. Cells were cultured for 10–20 days. Colonies were fixed in 4% formaldehyde and stained with 0.5% crystal violet. Cells were photographed using a Leica microscope. Colonies were then destained using 10% acetic acid and crystal violet staining was quantified by measuring absorbance at 595 nm using a Spectramax M5 instrument (Molecular Devices).

Immunofluorescence

For immunofluorescence assays, 200,000 cells were plated on SecureSlip silicone supported coverglasses (Sigma Aldrich) in 6-well plates that had been pre-coated for 60 minutes with 0.01 mg/mL human fibronectin (Calbiochem) in PBS. The following day, cells were fixed in 4% paraformaldehyde (PFA) diluted in PBS for 15 minutes at room temperature. Cells were permeabilized with 0.2% Triton X-100 in PBS for 10 minutes. Blocking was performed in 2.5% normal goat serum blocking solution (Vector Laboratories). Cells were incubated in primary MAGOH/MAGOHB antibody (Abcam, ab38768, rabbit, 1:200) for 1 hr at RT. A Cy-3 conjugated anti-rabbit secondary (Abcam, ab97075, 1:200) and DAPI (1:1000, Life Technologies) were then used for 1 hr at RT. Cells were mounted and imaged on an Axio Observer fluorescent microscope (Zeiss) using AxioVision software (Zeiss). Nuclear/Cytoplasmic ratio was quantified by Image J. Nuclear outlines were determined based on DAPI signal. Cytoplasmic signal was defined as signal in the whole cell minus signal within the nuclear area.

Mouse Experiments

Studies involving mice were approved by the MIT Committee on Animal Care. Mouse strain used as NCR-nude (Charles River Laboratories), female, 4–5 weeks of age. For inducible shRNA xenograft experiments, NCR-nude mice were subcutaneously injected into bilateral flanks with 3.5×10^6 H1437 cells transduced with lentivirus expressing a doxycycline-inducible shRNA against MAGOHB (MAGOHB-sh2). Cells were resuspended in 100 μL of 30% matrigel in PBS. At 7 days post-injection, after tumor implantation, mice were randomized to match tumor size between two groups, and one group was started on a diet containing 200 mg doxycycline/kg (Bio-Serv, Flemington, NJ). Tumor volumes were measured twice weekly using a digital caliper.

For tumor-penetrating nanocomplex (TPNC) experiments, xenografts were produced as above using 2.5×10^6 cells per tumor. TPNCs were prepared by complexing siRNA with

tandem peptide at a 1:20 (LyP1) or 1:15(iRGD) molar ratio (siRNA:peptide) in water. For intratumoral injections, 0.2nmol siRNA were injected every 1–2 days in 20uL of TPNC solution. Myr-TP-LyP1 (myr-GWTLNSAGYLLGKINLKALAALAKKILGGGG-K(5TAMRA)-CGNKRTRGC (C-C bridge)) and Palm-TP-iRGD (palm-GWTLNSAGYLLGKINLKALAALAKKILGGGG-CRGDKGPDC (C-C bridge)) were synthesized by CPC Scientific. siRNAs were purchased from GE Dharmacon. *MAGOHB* siRNA target sequence was as in *MAGOHB*-sh2. *IPO13* siRNA target sequence as in *IPO13*-sh3 (see Supplementary Table 14 for sequences). Surface expression of p32 or NRP-1 were evaluated by flow cytometry using anti-p32 antibody at 1:1000 dilution (AB2991, EMD Millipore) or anti-alpha V beta 3-PE conjugated antibody at 1:100 dilution (FAB3050P, R&D Systems) or anti-Neuropilin1 antibody at 1:1000 dilution (AB9600, Millipore), or matched isotype control, and visualized with AlexaFluor 647-labeled secondary antibody (p32 and NRP-1) or conjugated PE (αvβ3).

Immunostaining was performed as previously described[42]. Briefly, six tumors from each condition (randomly selected) were extracted and fixed in 10% formalin overnight and stored at 4˚C before being embedded in paraffin, sectioned, and stained. Tumor sections were stained with primary antibody to Cleaved Caspase-3 (Asp175) (5A1E) Rabbit mAb #9664 (Cell Signaling Technology, 1:1000) and HRP-conjugated anti-rabbit secondary antibody (RABBIT-ON-RODENT HRP-POLYMER from BioCare Medical, Cat #RMR622) on a ThermoScientific IHC Autostainer 360 and visualized with DAB chromogen. For cleaved caspase-3 quantification, fractional of cross-sectional area staining positive for cleaved caspase-3 was quantified in the six randomly selected tumors from each group that were stained, using ImageJ.

Immunoprecipitation and Mass Spectrometry

Immunoprecipitation.—For immunoprecipitation experiments, 293T cells were either untransduced (control) or transduced with pLX304-Blast-V5 (Addgene #25890) expressing MAGOH-V5 or MAGOHB-V5. Following antibiotic selection to derive stably transduced cell populations, immunoprecipitation was carried out using the Pierce Class Magnetic IP Kit (#88804) and anti-V5 magnetic beads (MBLI #M167–11) using a starting amount of 2 mg protein and 50 µl beads. Lysis buffer was pH 7.4, 0.025M Tris, 0.15M NaCl, 0.001M EDTA, 1% NP40, 5% glycerol. Immunoprecipitation was carried out overnight at 4^{o}C. Samples were washed twice in sample buffer, followed by twice in PBS prior to mass spectrometry. Efficient immunoprecipitation was confirmed by western blotting prior to proceeding with mass spectrometry.

Protein digestion and labeling with TMT isobaric mass tags.—The beads from immunopurification were washed once with IP lysis buffer, then three times with PBS, the four different lysates of each replicate were resuspended in 90 uL digestion buffer (2M Urea, 50 mM Tris HCl), 2 ug of sequencing grade trypsin added, 1 hour shaking at 700rpm. The supernatant was removed and placed in a fresh tube. The beads were then washed twice with 50 uL digestion buffer and combined with the supernatant. The combined supernatants were reduced (2 uL 500 mM DTT, 30 minutes, RT), alkylated (4 uL 500 mM IAA, 45 minutes,

dark) and a longer overnight digestion performed: 2 ug (4 uL) trypsin, shake o/n, The samples were then quenched with 20 uL 10% FA and desalted on 10 mg Oasis cartridges.

Desalted peptides were labeled with TMT6 reagents lot QD218427 (Thermo Fisher Scientific) according to the following: 126, NoBaitCntlRep1; 127, NoBaitCntlRep2; 128, MAGOHRep1; 129, MAGOHRep2; 130, MAGOHBRep1; 131, MAGOHBRep2. Peptides were dissolved in 25uL of fresh 100mM HEPES buffer. The labeling reagent was resuspended in 42 ul of acetonitrile and 10 ul added to each sample as described below. After 1 h incubation the reaction was stopped with 8uL of 5mM Hydroxylamine.

Protein identification with nanoLC–MS system.—Reconstituted peptides were separated on an online nanoflow EASY-nLC 1000 UHPLC system (Thermo Fisher Scientific) and analyzed on a benchtop Orbitrap Q Exactive plus mass spectrometer (Thermo Fisher Scientific). The peptide samples were injected onto a capillary column (Picofrit with 10 μm tip opening / 75 μm diameter, New Objective, PF360–75-10-N-5) packed in-house with 20 cm C18 silica material (1.9 μm ReproSil-Pur C18-AQ medium, Dr. Maisch GmbH) and heated to 50 °C in column heater sleeves (Phoenix-ST) to reduce backpressure during UHPLC separation. Injected peptides were separated at a flow rate of 200 nL/min with a linear 230 min gradient from 100% solvent A (3% acetonitrile, 0.1% formic acid) to 30% solvent B (90% acetonitrile, 0.1% formic acid), followed by a linear 9 min gradient from 30% solvent B to 60% solvent B and a 1 min ramp to 90%B. Each sample was run for 260 min, including sample loading and column equilibration times. The Q Exactive instrument was operated in the data-dependent mode acquiring HCD MS/MS scans (R=17,500) after each MS1 scan (R=70,000) on the 12 top most abundant ions using an MS1 ion target of 3×10^6 ions and an MS2 target of 5×10^4 ions. The maximum ion time utilized for the MS/MS scans was 120 ms; the HCD-normalized collision energy was set to 27; the dynamic exclusion time was set to 20s, and the peptide match and isotope exclusion functions were enabled.

Database search and data processing.—All mass spectra were processed using the Spectrum Mill software package v6.0 pre-release (Agilent Technologies) which includes modules developed by us for TMT6 -based quantification. For peptide identification MS/MS spectra were searched against human Uniprot database to which a set of common laboratory contaminant proteins was appended. Search parameters included: ESI-QEXACTIVE-HCD scoring parameters, trypsin enzyme specificity with a maximum of two missed cleavages, 40% minimum matched peak intensity, +/− 20 ppm precursor mass tolerance, +/− 20 ppm product mass tolerance, and carbamidomethylation of cysteines and TMT6 labeling of lysines and peptide n-termini as fixed modifications. Allowed variable modifications were oxidation of methionine, N-terminal acetylation, Pyroglutamic acid (N-termQ),Deamidated (N),Pyro Carbamidomethyl Cys (N-termC),with a precursor MH+ shift range of −18 to 64 Da. Identities interpreted for individual spectra were automatically designated as valid by optimizing score and delta rank1-rank2 score thresholds separately for each precursor charge state in each LC-MS/MS while allowing a maximum target-decoy-based false-discovery rate (FDR) of 1.0% at the spectrum level.

Analysis.—The expectation maximization algorithm[43] was applied to the results of the peptide report (the in-house written bash script is available on the Lage Lab Resources Site (see URLs) and the peptide report can be found in the supplementary materials). The list of most likely observed proteins was generated for each channel of the MS experiment based on Swiss-Prot and TrEMBLE databases of protein sequences[44]. Next, ratios of intensities between channels were calculated and median normalized. Resulting data were analyzed and visualized using R. Statistical analyses were performed via moderated t-test from R package limma[45] to estimate p values for each protein and the false discovery rate corrections (FDR) were applied to account for multiple hypothesis testing. Plots were created using in-house written R scripts and gplot2[46]. RNA-binding and S ribosomal protein families were taken from HUGO Gene Nomenclature Committee (see URLs). Proteins previously reported to be EJC/NMD complex members[7] were annotated as such.

Analysis of Cancer Cell Lines

Copy Number Analysis—Details regarding arm-level copy number calling are as described in Taylor et al. (2018)[47]. Briefly, to determine arm-level events (i.e. 1p or 12p deletion status) in TCGA and CCLE samples, the ABSOLUTE algorithm[48] was used to determine the likeliest ploidy and absolute total copy number of each genomic segment. Segments were called as amplified, deleted or copy neutral based on copy number with reference to integer-rounded ploidy. Arm or chromosome level amplification/deletion status was then determined from segment data as described in Taylor et al. (2018)[47].CCLE cell lines were fit to clusters within their corresponding TCGA tumor type to generate cell line-specific, arm-specific calls[49]. For CCLE data, ABSOLUTE algorithm was run on the CCLE Affymetrix SNP6.0 array data as previously reported[50]. For analysis of arm-level and focal copy number event in TCGA data sets, 1p deletion status was determined as described above. Hemizygous *MAGOH* deletion was defined as the loss of one or more copies of the MAGOH gene (e.g. ploidy – MAGOH copy number >= 1) using rounded tumor ploidy and *MAOGH* copy number calculated from the ABSOLUTE algorithm.

Genome-scale shRNA and CRISPR Screening Data Analysis—Genome-wide shRNA screening on 501 cell lines was performed as described[5]. The DEMETER method, which summarizes multiple shRNAs targeting a gene into a gene-level dependency score, was used to quantify gene dependency in 17,098 unique genes[5]. The differential dependency set of 6,305 genes, and the 6σ dependency set of 769 genes, as defined previously[5], were used for all downstream analysis. These sets represent the genes with the most significant differential dependency across cell lines and were selected based on the following criteria: 1) for each gene, there is at least one cell line with a dependency score that is two (differential set) or six (6σ set) standard deviations from the mean of scores from all genes and all cell lines, and 2) expression of the gene in the most dependent cell line is above –2 log 2 RPKM.

To identify synthetic lethal relationships linked to loss of a paralog, a query was performed for each of 17,670 genes using EnsemblCompara[51] via the R interface to BioMart[52] to obtain a list of paralogs and their pairwise sequence identity. Pearson correlations of RNAseq expression values between genes in each paralog pair indicate that co-expression is limited until DNA sequence identity exceeds 25% (Supplementary Figure 1). To increase the

likelihood that the gene pairs function redundantly, pairs with less than 25% sequence identity were removed. An additional 35 genes were removed for having duplicate DEMETER scores (caused by non-unique hairpins), resulting in 3,403 genes in the DEMETER dataset with at least one paralog. Differential dependency for each of these genes was tested by grouping cell lines based on loss of the gene's paralogs and performing a two-class comparison of the DEMETER scores using empirical Bayes moderated t-statistics implemented by the R package 'limma'[45]. The binary classification of paralog loss used to group the cell lines was determined by a logic combination of the RNAseq gene expression (GE), protein abundance (RPPA), relative copy number (CN), methylation fraction (RRBS) and mutation status (WES,WGS,RNAseq). The GE, RPPA, CN, and RRBS datasets are z-scored per gene so loss of a gene is defined as having a 6σ decrease in GE or RPPA, or no GE at all (less than −3 log2 TPM), 2σ decrease in CN, 6σ increase in RRBS, or a deleterious mutation (predicted by frameshift indel or nonsense SNV). Genes are labeled 'symmetric' if loss of either gene in a pair is significantly associated with a selective dependency on its respective paralog gene with FDR < .05.

The synthetic lethal paralog analysis was repeated using the Achilles CRISPR dataset[15] consisting of 341 whole genome CRISPR/cas9 knockout screens corrected for copy number effects (one cell line, PK59, was removed from the prior set of 342 as it failed fingerprinting). Genes with variance in essentiality below .01 across the 341 cell lines were removed to reduce false positives, leaving 6,535 genes for paralog dependency analysis. The definition of gene loss as well as method for determining significance of differential dependency among each paralog pair is identical to the analysis using DEMETER data.

For analysis of gene dependencies correlated with *MAGOH* deletion (Fig 1), the PARIS algorithm was run as a GenePattern module (https://www.broadinstitute.org/achilles/resources). *MAGOH*-deletion status was determined by the ABSOLUTE algorithm[48] as described above. Cell lines for which *MAGOH* absolute copy number was less than the cell line's ploidy were considered deleted. Based on available ABSOLUTE calls, 191 lines were considered deleted and 807 lines were considered non-deleted. In total, both absolute copy number and filtered DEMETER gene-score data was available for 445 overlap cell lines.

For geneset enrichment analysis on gene dependencies correlated with *MAGOH* deletion, the PARIS algorithm was first run using continuous copy number data on CCLE cell lines generated using SNP arrays, as previously reported[50] to generate a ranked list of gene dependencies correlated with *MAGOH* copy number. RNMI metric score for each gene was then used as input for preranked geneset enrichment analysis[53], which was run as a GenePattern module using default parameters against the following genesets:REACTOME_NONSENSE_MEDIATED_DECAY_ENHANCED_BY_THE_EXON_JUNCTION_COMPLEX and GO_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS _NONSENSE_MEDIATED_DECAY

For analysis of correlated gene dependency profiles, Pearson correlations of DEMETER gene dependency scores were computed across cell lines (N=501) for all pairs of genes that share overlap in cell lines (N=6,300). Correlation coefficients were converted to standard

scores across the full correlation matrix before evaluating the specific MAGOH and MAGOHB correlation profiles.

RNA-Seq Analysis—RNA-Seq libraries were prepared using the Illumina strand-specific mRNA seq library prep kit (Illumina) followed by paired-end 75 bp sequencing on a NextSeq (>400M reads per run; >33M reads/sample). Transcript levels were quantified with kallisto[19] (version 0.43.0, options: –rf-stranded, –b 30) using the GRCh38 transcriptome (ENSEMBL cDNA, release 87)[54]. Differential expression analysis was performed with sleuth[55] (version 0.28.1). Differential expression was quantified based on the "beta" value, a bias estimator used by kallisto[19] analogous to fold-change. Significant upregulation cut-offs were b>1, q<0.01; downregulation b < −1. Gene Ontology (GO)[56] term enrichment analysis was carried out using PANTHER[57]. Overrepresentation Test (release 20160715) using the GO database (release 2017–01-26), accessible via the GO website (www.geneontology.org, last accessed 2017–01-31). Transcript biotypes were obtained from the ENSEMBL database (release 87)[54]. For analysis of differential alternative splicing events, reads were aligned with HiSat2 (v2.0.4, --rna-strandness RF option)[58] using the prebuilt index Ensembl GRCh38 (genome_tran), and splicing events were quantified using rMATS v3.2.5[59]. For increased stringency, rMATS output was filtered based on read support (sum of inclusion/exclusion reads >= 10 in both samples), FDR (<0.05), and inclusion level difference (|ILD| >0.1). Sashimi plots were plotted using rMats2Sashimiplot (see URLs) in grouping mode.

For estimation of protein-level effects from RNA-Seq data, peptide sequences for each transcript were obtained using ENSEMBL biomart (see URLs), accessed through the R package biomaRt (version 2.26.1) [52,60]. For each gene, expected peptide expression was then estimated by summing over TPM values for all transcripts that encode peptides of the same length.

REVEALER analysis—To identify associations between *MAGOH/B* dependency and copy number/expression features of EJC/splicing related genes, *MAGOH* or *MAGOHB* dependency scores across screened cell lines[5] were correlated to copy number or expression features[50] in EJC/splicing-related genes using the previously described method based on estimating the Information Coefficient[61]. For radial plots, the top scoring 50 features (for copy number) or top scoring 16 features (for expression) were plotted. A list of EJC/splicing-related genes used for this analysis was compiled by combining EJC/NMD genes in MSigDB (Reactome geneset #M1067) and splicing factors described to be implicated in oncogenesis[62].

Statistics

No statistical methods were used to pre-determine sample size. Investigators were not blinded to allocation for experiments. Statistical tests applied, p-values, and sample size are as listed in figure captions. For *in vitro* experiments, number of biologically independent replicates is as listed in figure captions. When two-sample Student's t-tests were applied to assess significance of experimental data, unequal variance parameter was used and p-values were calculated using Microsoft Excel (function t.test; heteroscedastic). Other statistical tests were performed using R (v. 3.4.1) or GraphPad Prism 7 software.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS:

S.R.V. was supported by a Young Investigator Award from the American Society of Clinical Oncology. This work was supported by a National Cancer Institute grant 1R35CA197568 and an American Cancer Society Research Professorship to M.M. P.T. was supported by NIH grants U01CA217885 and R01HG009285. W.C.H. was supported by U01CA176058. C.B. and S.N.B. were supported by a Koch Institute Support Grant (P30-CA14051) from the National Cancer Institute (Swanson Biotechnology Center) and a Core Center Grant (P30-ES002109) from the National Institute of Environmental Health Sciences, and the Ludwig Center for Molecular Oncology. C.G.B. was supported by the National Science Foundation Graduate Research Fellowship Program. S.N.B. is a Howard Hughes Medical Institute Investigator. P.S.C. was supported by an NIH Pathway to Independence Award (K99 CA208028). E.M. would like to thank Heiko Horn for help with the expectation maximization algorithm. The authors thank the Koch Institute Swanson Biotechnology Center for technical support, specifically Kathleen Cormier in the Hope Babette Tang Histology Facility.

REFERENCES (FOR MAIN TEXT)

1. Diss G et al. Gene duplication can impart fragility, not robustness, in the yeast protein interaction network. Science 355, 630–634 (2017). [PubMed: 28183979]
2. Muller FL, Aquilanti EA & DePinho RA Collateral Lethality: A new therapeutic strategy in oncology. Trends Cancer 1, 161–173 (2015). [PubMed: 26870836]
3. Frei E Gene deletion: a new target for cancer chemotherapy. The Lancet 342, 662–664 (1993).
4. McDonald ER et al. Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. Cell 170, 577–592.e10 (2017). [PubMed: 28753431]
5. Tsherniak A et al. Defining a Cancer Dependency Map. Cell 170, 564–576.e16 (2017). [PubMed: 28753430]
6. Boehm V & Gehring NH Exon Junction Complexes: Supervising the Gene Expression Assembly Line. *Trends Genet.* TIG 32, 724–735 (2016). [PubMed: 27667727]
7. Chang Y-F, Imam JS & Wilkinson MF The Nonsense-Mediated Decay RNA Surveillance Pathway. Annu. Rev. Biochem. 76, 51–74 (2007). [PubMed: 17352659]
8. Singh KK, Wachsmuth L, Kulozik AE & Gehring NH Two mammalian MAGOH genes contribute to exon junction complex composition and nonsense-mediated decay. RNA Biol. 10, 1291–1298 (2013). [PubMed: 23917022]
9. Mingot J-M, Kostka S, Kraft R, Hartmann E & Görlich D Importin 13: a novel mediator of nuclear import and export. EMBO J. 20, 3685–3694 (2001). [PubMed: 11447110]
10. Helming KC et al. ARID1B is a specific vulnerability in ARID1A-mutant cancers. Nat. Med. 20, 251–254 (2014). [PubMed: 24562383]
11. Hoffman GR et al. Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. *Proc. Natl.* Acad. Sci. 111, 3128–3133 (2014). [PubMed: 24520176]
12. D'Antonio M et al. Recessive cancer genes engage in negative genetic interactions with their functional paralogs. Cell Rep. 5, 1519–1526 (2013). [PubMed: 24360954]
13. Dey P et al. Genomic deletion of malic enzyme 2 confers collateral lethality in pancreatic cancer. Nature 542, 119–123 (2017). [PubMed: 28099419]
14. Cowley GS et al. Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies. Sci. Data 1, 140035 (2014). [PubMed: 25984343]
15. Meyers R Computational correction of copy-number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells *Nature Genetics* In Press, (2017).
16. Veitia RA Gene Duplicates: Agents of Robustness or Fragility? Trends Genet. 33, 377–379 (2017). [PubMed: 28434610]

17. Bono F & Gehring NH Assembly, disassembly and recycling. RNA Biol. 8, 24–29 (2011). [PubMed: 21289489]
18. Caron H et al. Allelic loss of chromosome 1p as a predictor of unfavorable outcome in patients with neuroblastoma. *N. Engl.* J. Med. 334, 225–230 (1996). [PubMed: 8531999]
19. Bray NL, Pimentel H, Melsted P & Pachter L Near-optimal probabilistic RNA-seq quantification. *Nat. Biotech*nol. 34, 525–527 (2016). [PubMed: 27043002]
20. Ni JZ et al. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev. 21, 708–718 (2007). [PubMed: 17369403]
21. Lareau LF, Inada M, Green RE, Wengrod JC & Brenner SE Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature 446, 926–929 (2007). [PubMed: 17361132]
22. Lau C-K, Diem MD, Dreyfuss G & Van Duyne GD Structure of the Y14-Magoh Core of the Exon Junction Complex. *Curr.* Biol. 13, 933–941 (2003). [PubMed: 12781131]
23. Ren Y et al. Targeted tumor-penetrating siRNA nanocomplexes for credentialing the ovarian cancer oncogene ID4. *Sci. Transl.* Med. 4, 147ra112 (2012).
24. Beroukhim R et al. The landscape of somatic copy-number alteration across human cancers. Nature 463, 899–905 (2010). [PubMed: 20164920]
25. Baudis M Genomic imbalances in 5918 malignant epithelial tumors: an explorative meta-analysis of chromosomal CGH data. BMC Cancer 7, 226 (2007). [PubMed: 18088415]
26. Nijhawan D et al. Cancer Vulnerabilities Unveiled by Genomic Loss. Cell 150, 842–854 (2012). [PubMed: 22901813]
27. Paolella BR et al. Copy-number and gene dependency analysis reveals partial copy loss of wild-type SF3B1 as a novel cancer vulnerability. eLife 6, (2017).
28. Lee SC-W et al. Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins. Nat. Med. 22, 672–678 (2016). [PubMed: 27135740]
29. Obeng EA et al. Physiologic Expression of Sf3b1(K700E) Causes Impaired Erythropoiesis, Aberrant Splicing, and Sensitivity to Therapeutic Spliceosome Modulation. Cancer Cell 30, 404–417 (2016). [PubMed: 27622333]
30. Lu G et al. The Myeloma Drug Lenalidomide Promotes the Cereblon-Dependent Destruction of Ikaros Proteins. Science 343, 305–309 (2014). [PubMed: 24292623]
31. Krönke J et al. Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. Science 343, 301–305 (2014). [PubMed: 24292625]
32. Hwang S-Y et al. Direct Targeting of β-Catenin by a Small Molecule Stimulates Proteasomal Degradation and Suppresses Oncogenic Wnt/β-Catenin Signaling. Cell Rep. 16, 28–36 (2016). [PubMed: 27320923]
33. Kerres N et al. Chemically Induced Degradation of the Oncogenic Transcription Factor BCL6. Cell Rep. 20, 2860–2875 (2017). [PubMed: 28930682]
34. Han T et al. Anticancer sulfonamides target splicing by inducing RBM39 degradation via recruitment to DCAF15. Science 356, (2017).
35. Soderholm JF et al. Importazole, a Small Molecule Inhibitor of the Transport Receptor Importin-β. ACS Chem. Biol. 6, 700–708 (2011). [PubMed: 21469738]
36. Hintersteiner M et al. Identification of a small molecule inhibitor of importin beta mediated nuclear import by confocal on-bead screening of tagged one-bead one-compound libraries. ACS Chem. Biol. 5, 967–979 (2010). [PubMed: 20677820]
37. Wagstaff KM, Sivakumaran H, Heaton SM, Harrich D & Jans DA Ivermectin is a specific inhibitor of importin α/β-mediated nuclear import able to inhibit replication of HIV-1 and dengue virus. Biochem. J. 443, 851–856 (2012). [PubMed: 22417684]

REFERENCES (METHODS ONLY)

38. Yang X et al. A public genome-scale lentiviral expression library of human ORFs. Nat. Methods 8, 659–661 (2011). [PubMed: 21706014]
39. Lippa MS et al. Expression of anti-apoptotic factors modulates Apo2L/TRAIL resistance in colon carcinoma cells. Apoptosis Int. J. Program. Cell Death 12, 1465–1478 (2007).
40. Brown CY et al. Robust, reversible gene knockdown using a single lentiviral short hairpin RNA vector. Hum. Gene Ther. 21, 1005–1017 (2010). [PubMed: 20615123]
41. Root DE, Hacohen N, Hahn WC, Lander ES & Sabatini DM Genome-scale loss-of-function screening with a lentiviral RNAi library. Nat. Methods 3, 715–719 (2006). [PubMed: 16929317]
42. Kwon EJ, Dudani JS & Bhatia SN Ultrasensitive tumour-penetrating nanosensors of protease activity. Nat. Biomed. Eng. 1, 0054 (2017). [PubMed: 28970963]
43. McLachlan GJ & Krishnan T The EM algorithm and extensions. (Wiley-Interscience, 2008).
44. The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45, D158–D169 (2017). [PubMed: 27899622]
45. Ritchie ME et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47–e47 (2015). [PubMed: 25605792]
46. Wickham H ggplot2. (Springer New York, 2009). doi:10.1007/978-0-387-98141-3
47. Taylor AM et al. Genomic and Functional Approaches to Understanding Cancer Aneuploidy. Cancer Cell (2018).
48. Carter SL et al. Absolute quantification of somatic DNA alterations in human cancer. Nat. Biotechnol. 30, 413–421 (2012). [PubMed: 22544022]
49. Pedregosa F et al. Scikit-learn: Machine Learning in Python. JMLR 12, 2825–2830 (2011).
50. Barretina J et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 483, 603–607 (2012). [PubMed: 22460905]
51. Vilella AJ et al. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Res. 19, 327–335 (2008). [PubMed: 19029536]
52. Durinck S, Spellman PT, Birney E & Huber W Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4, 1184–1191 (2009). [PubMed: 19617889]
53. Subramanian A et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. 102, 15545–15550 (2005). [PubMed: 16199517]
54. Aken BL et al. Ensembl 2017. Nucleic Acids Res. 45, D635–D642 (2017). [PubMed: 27899575]
55. Pimentel HJ, Bray N, Puente S, Melsted P & Pachter L Differential analysis of RNA-Seq incorporating quantification uncertainty. *bioRxiv* 058164 (2016). doi:10.1101/058164
56. The Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res. 43, D1049–D1056 (2015). [PubMed: 25428369]
57. Mi H, Muruganujan A, Casagrande JT & Thomas PD Large-scale gene function analysis with the PANTHER classification system. Nat. Protoc. 8, 1551–1566 (2013). [PubMed: 23868073]
58. Kim D, Langmead B & Salzberg SL HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360 (2015). [PubMed: 25751142]
59. Shen S et al. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc. Natl. Acad. Sci. 111, E5593–E5601 (2014). [PubMed: 25480548]
60. Durinck S et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics 21, 3439–3440 (2005). [PubMed: 16082012]
61. Kim JW et al. Characterizing genomic alterations in cancer by complementary functional associations. Nat. Biotechnol. 34, 539–546 (2016). [PubMed: 27088724]
62. Sveen A, Kilpinen S, Ruusulehto A, Lothe RA & Skotheim RI Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. Oncogene (2015). doi:10.1038/onc.2015.318

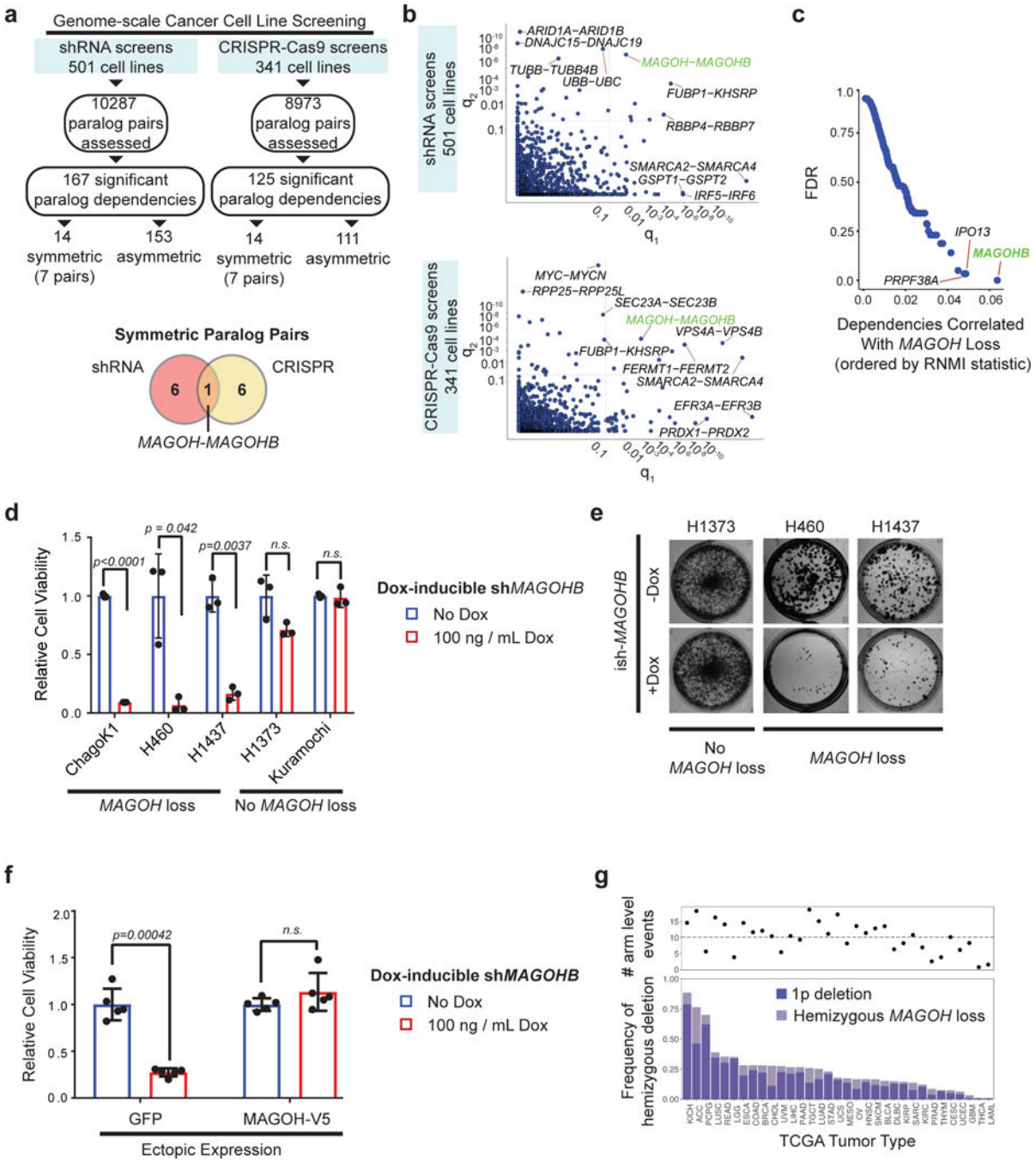

Figure 1. Hemizygous MAGOH deletion confers MAGOHB dependency.

(a) Analysis of paralog dependencies in genome-scale screening of cancer cell lines (shRNA, 501 lines; CRISPR-Cas9, 341 lines).

(b) q-value:q-value plot showing significance of pairwise correlation between a gene's dependency score and inactivation of its paralog. q-value 1, significance for dependency on the paralog labeled first with inactivation of the paralog labeled second. q-value 2, significance for dependency on the paralog labeled second with inactivation of the paralog labeled first. "Symmetric" paralogs are in upper right quadrant ($q1 < 0.05$ and $q2 < 0.05$).

Plots show n=1970 paralog pairs for shRNA data and n=1593 pairs for CRISPR data. One-sided p-value from two class comparison was calculated via moderated t-statistic and adjusted for multiple comparisons using the Benjamini-Hochberg false-discovery rate (FDR).

(c) Probability Analysis by Ranked Information Score (PARIS) analysis to identify gene dependencies correlated with hemizygous *MAGOH* loss. Mutual information metric (RNMI) is plotted against FDR for gene dependencies positively correlated with *MAGOH* deletion.

(d) Frequency of hemizygous *MAGOH* deletion across TCGA cohorts. Total frequency of *MAGOH* loss is indicated by a light blue bar; frequency of *MAGOH* loss occurring as a result of chromosome 1p deletion is indicated by a dark blue bar. Top panel shows total number of arm level copy number events in each tumor type.

(e) Cell viability in cell lines with (left) and without (right) hemizygous *MAGOH* loss upon *MAGOHB* suppression using a doxycycline-inducible shRNA against *MAGOHB*. Error bars show mean +/− s.d., n=3 replicates from a representative experiment repeated at least twice in each cell line; p-value by two-tailed, two-sample t-test.

(f) Colony formation in cell lines with (H1437, H460) or without (H1373) hemizygous *MAGOH* loss upon *MAGOHB* suppression using a doxycycline-inducible shRNA against *MAGOHB*. Photographs show representative wells from an experiment conducted in triplicate (quantification in Supplementary Figure 4); experiment was repeated at least twice in each cell line

(g) Cell viability measured upon *MAGOHB* knockdown in ChagoK1 cells with or without reconstitution of MAGOH-V5. Error bars show mean +/− s.d., n=5 replicates from a representative experiment repeated at least twice; p-value by two-tailed, two-sample t-test

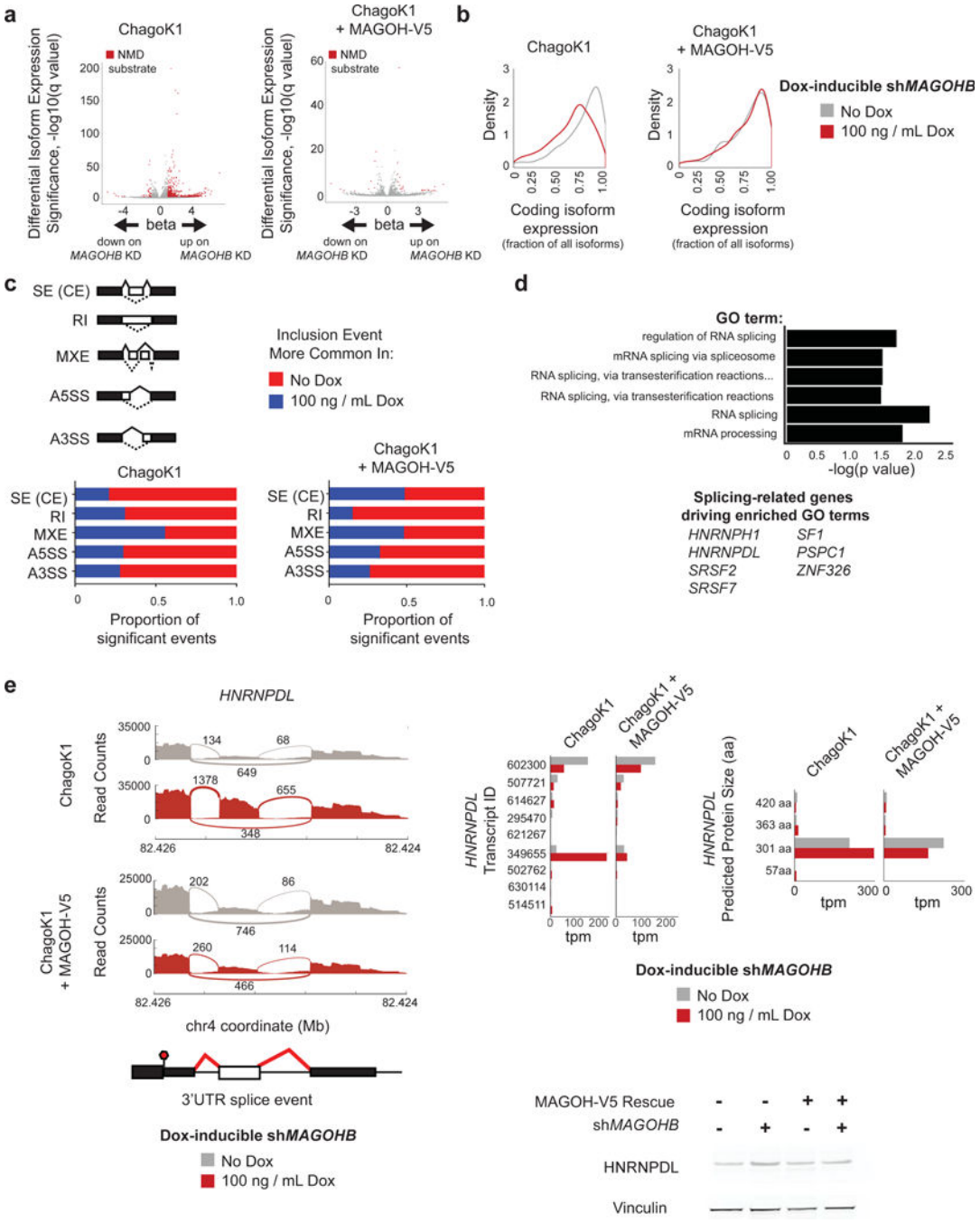

Figure 2. RNA splicing is globally altered upon MAGOHB suppression in cells with MAGOH loss, leading to the upregulation of NMD substrates.

(a) Differentially expressed transcripts in ChagoK1 cells (**left**) and in *MAGOH*-reconstituted ChagoK1 cells upon *MAGOHB* knockdown (**right**). Transcripts annotated as NMD substrates shown in red. Significance determined by a Wald test and adjusted for multiple comparisons using the Benjamini-Hochberg FDR. n=3 replicates were used in all conditions. **(b)** Density distribution of proportional expression levels among coding isoforms corresponding to genes whose NMD isoforms are upregulated upon *MAGOHB* knockdown

in ChagoK1 cells (**left**) or *MAGOH*-reconstituted ChagoK1 cells (**right**). X-axis shows expression level of coding isoform(s) proportional to all expressed transcripts for a given gene; Y-axis shows density.

(c) Global changes in patterns of splice site usage upon *MAGOHB* knockdown in ChagoK1 cells (**left**) or MAGOH-V5 reconstituted ChagoK1 cells (**right**). Splice event classes are shown in the schematic; solid lines denote "inclusion" event and dotted lines denote the alternative event for each class. Bar graphs denote the proportion of significant differentially spliced events that show greater inclusion in either the absence (red) or presence (blue) of *MAGOHB* knockdown in either ChagoK1 cells (left) or MAGOH-V5 reconstituted ChagoK1 cells (right).

(d) Significantly enriched Gene Ontology classes for genes (n=17) that show upregulation of NMD isoform(s) and concomitant downregulation of coding isoform(s). Significance was determined by a binomial test and adjusted for multiple comparisons using the Bonferroni correction.

(e) left) Sashimi plots around an activated exon within the 3'UTR of the *HNRNPDL* gene (inclusion of which creates an NMD-substrate transcript) in either the absence or presence of *MAGOHB* knockdown in either ChagoK1 cells or MAGOH-V5 reconstituted ChagoK1 cells. Numbers reflect junction spanning reads averaged over three replicates for each condition.

right, top) Left panel, Transcript abundances for various isoforms of the *HNRNPDL* gene in either the absence (grey) or presence (red) of *MAGOHB* knockdown in either ChagoK1 cells or MAGOH-V5 reconstituted ChagoK1 cells. Right panel, isoform abundances grouped by predicted coding protein length in each condition.

right, bottom) Western blot showing increased HNRNPDL protein levels upon *MAGOHB* knockdown in ChagoK1 cells but not MAGOH-V5 reconstituted ChagoK1 cells. Representative of a similar experiment repeated three times.

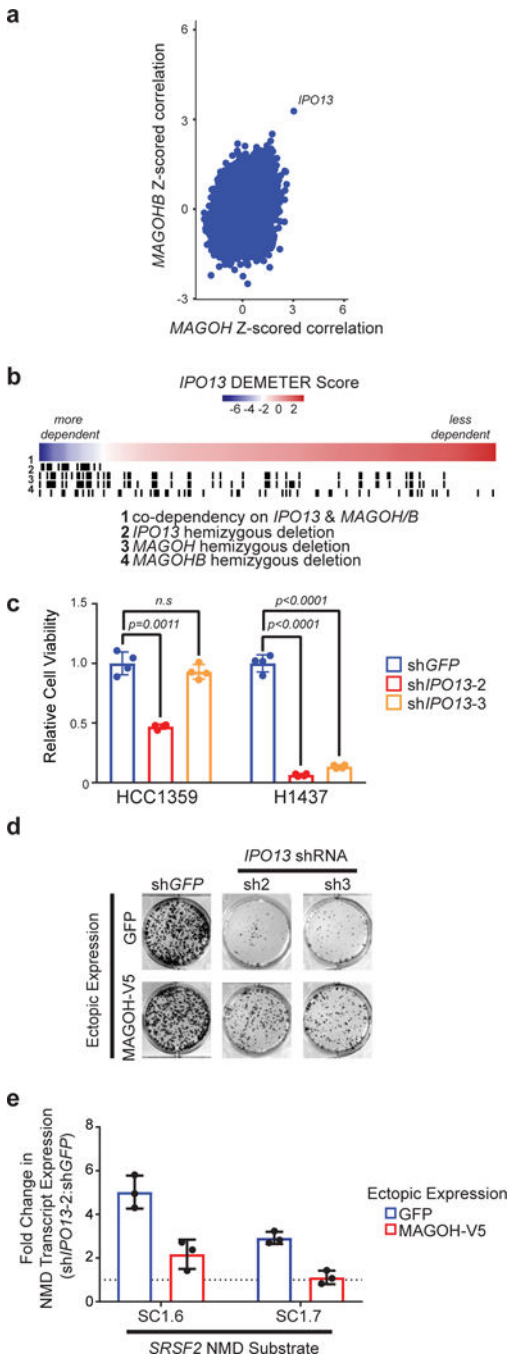

Figure 3. *I*PO13 dependency is correlated with MAGOH and MAGOHB dependencies and is rescued by MAGOH reconstitution.

(a) Plot of gene dependencies (n=6300) correlated to *MAGOH* dependency vs. those correlated to *MAGOHB* dependency. Axes reflect Z-scored Pearson correlation of each dependency to either *MAGOH* dependency (X-axis) or *MAGOHB* dependency (Y-axis). *MAGOH* and *MAGOHB* self-correlations are not shown.

(b) Heatmap of *IPO13* dependency scores across 243 screened cell lines. Black bars denote cell lines that share dependency on *IPO13* and either *MAGOH,MAGOHB,* or both (1); cell lines that carry a hemizygous deletion in *MAGOH* (2); *IPO13* (3); *MAGOHB* (4).
(c) Cell viability measured upon shRNA-mediated *IPO13* knockdown in cell line without (HCC1359, left) and with (H1437, right) *MAGOH* loss. Error bars show mean +/− s.d, n=4 replicates per cell line; p-value by two-tailed, two-sample t-test.
(d) Colony formation in *MAGOH*-deleted H460 cells upon *IPO13* knockdown in either the absence (top) or presence (bottom) of MAGOH-V5 reconstitution. Photographs show representative wells from an experiment conducted in triplicate (quantification in Supplementary Figure 11); experiment was repeated three times.
(e) Fold change in expression of the NMD substrates SC1.6 and SC1.7 of the *SRSF2* gene in H460 cells upon *IPO13* knockdown (IPO13-sh2) in either the presence (red) or absence (blue) of MAGOH-V5 reconstitution. Data normalized to expression in the s*hGFP* condition. Error bars show mean +/− s.d., n=3 technical replicates per condition.

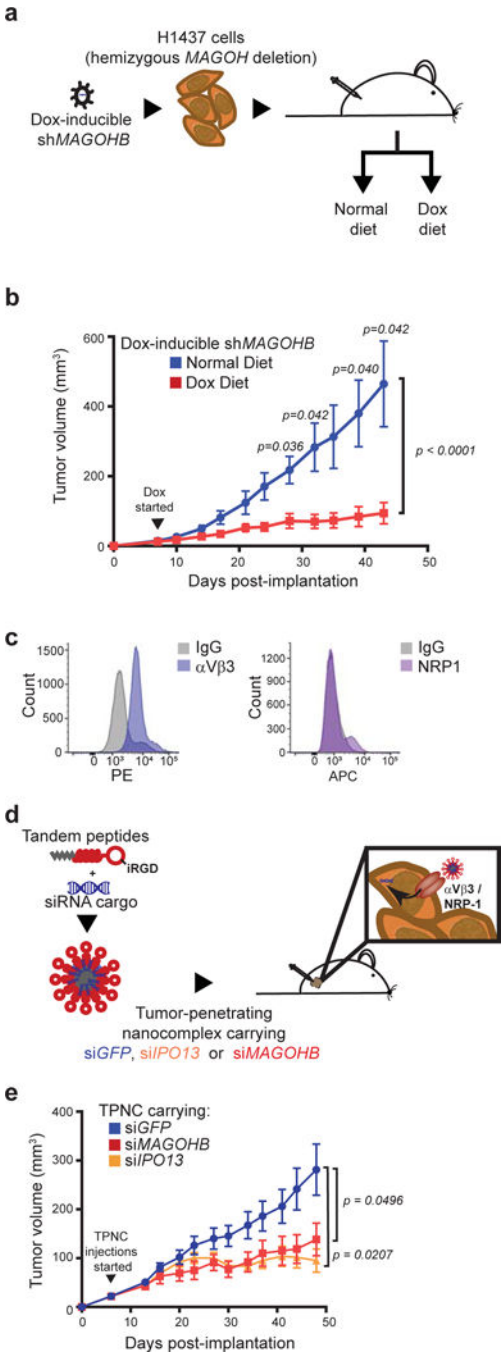

Figure 4. MAGOHB and IPO13 are in vivo dependencies in MAGOH-deleted xenografts.

(a,b) Schematic **(a)** and growth curves **(b)** of H1437 xenografts in nude mice upon *MAGOHB* suppression. H1437 cells were transduced with lentivirus expressing a doxycycline-inducible shRNA against *MAGOHB* and injected into the flanks of nude mice. Once palpable tumors had formed (7d), mice were randomized to either normal chow or chow supplemented with doxycycline. Tumor volume over time is plotted in each arm. Lines show mean +/− s.e.m for n=6 tumors per arm. p-values listed for significant (<0.05) time

points by two-tailed, two-sample t-test; p also < 0.0001 by two-way ANOVA between +Dox and –Dox curves.

(c) Surface expression for NRP-1 and αVβ3 (co-receptors for iRGD-containing nanocomplexes) in H1437 cells, as assessed by flow cytometry. Repeated twice with similar results.

(d,e) Schematic **(d)** and growth curves **(e)** of H1437 xenografts in nude mice upon *MAGOHB* or *IPO13* suppression using an siRNA-carrying tumor-penetrating nanocomplex (TPNC). Following palpable tumor formation, mice received intra-tumoral injections of TPNC containing either siRNA against *GFP* (control) *MAGOHB*, or *IPO13*. Lines show mean +/− s.e.m. for n=10 tumors per arm; p-value by two-way ANOVA.